AFRL-RX-WP-JA-2016-0306

STRUCTURAL VARIATION OF ALPHA-SYNUCLEIN WITH TEMPERATURE BY A COARSE-GRAINED APPROACH WITH KNOWLEDGE-BASED INTERACTIONS (POSTPRINT)

Peter Mirau and B.L. Farmer

AFRL/RX

R.B. Pandey

University of Mississippi

7 October 2015
Interim Report

(STINFO COPY)

AIR FORCE RESEARCH LABORATORY
MATERIALS AND MANUFACTURING DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7750
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
7 October 2015	Interim	8 September 2014 – 7 September 2015

4. TITLE AND SUBTITLE
STRUCTURAL VARIATION OF ALPHA-SYNUCLEIN WITH TEMPERATURE BY A COARSE-GRAINED APPROACH WITH KNOWLEDGE-BASED INTERACTIONS (POSTPRINT)

5a. CONTRACT NUMBER
FA8650-15-D-5405-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62102F

6. AUTHOR(S)
1) Peter Mirau and B.L. Farmer – AFRL/RX
2) R.B. Pandey – University of Mississippi

5d. PROJECT NUMBER
4348

5e. TASK NUMBER   0001

5f. WORK UNIT NUMBER
X0YC

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
1) AFRL/RX
Wright-Patterson AFB, OH 45433

2) University of Mississippi
118 College Drive
Hattiesburg, MS 38677

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Materials and Manufacturing Directorate
Wright-Patterson Air Force Base, OH 45433-7750
Air Force Materiel Command
United States Air Force

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RXAS

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RX-WP-JA-2016-0306

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
PA Case Number: 88ABW-2015-4825; Clearance Date: 7 Oct 2015. This document contains color. Journal article published in AIP Advances, Vol. 5, No. 9, 1 Jul 2015. © 2015 AIP Publishing. The U.S. Government is joint author of the work and has the right to use, modify, reproduce, release, perform, display, or disclose the work.
The final publication is available at doi: http://dx.doi.org/10.1063/1.4927544

14. ABSTRACT *(Maximum 200 words)*
Despite enormous efforts, our understanding the structure and dynamics of α-synuclein (ASN), a disordered protein (that plays a key role in neurodegenerative disease) is far from complete. In order to better understand sequence-structure-property relationships in α-SYNUCLEIN we have developed a coarse-grained model using knowledge-based residue-residue interactions and used it to study the structure of free ASN as a function of temperature (T) with a large-scale Monte Carlo simulation. Snapshots of the simulation and contour contact maps show changes in structure formation due to self-assembly as a function of temperature. Variations in the residue mobility profiles reveal clear distinction among three segments along the protein sequence. The N-terminal (1-60) and C-terminal (96-140) regions contain the least mobile residues, which are separated by the higher mobility non-amyloid component (NAC) (61-95). Our analysis of the intra-protein contact profile shows a higher frequency of residue aggregation (clumping) in the N-terminal region relative to that in the C-terminal region, with little or no aggregation in the NAC region. The radius of gyration (Rg) of ASN decays monotonically with decreasing the temperature, consistent with the finding of Allison et al. (JACS, 2009).

15. SUBJECT TERMS
α-synuclein (ASN); Monte Carlo simulation; protein sequence; N-terminal; C-terminal; non-amyloid component (NAC)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	SAR	12	Lawrence Brott
					19b. TELEPHONE NUMBER *(Include Area Code)* (937) 255-9157

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

Structural variation of alpha-synuclein with temperature by a coarse-grained approach with knowledge-based interactions

Peter Mirau,[1] B. L. Farmer,[1] and R. B. Pandey[2]
[1]*Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, OH 45433, USA*
[2]*Department of Physics and Astronomy, University of Southern Mississippi, Hattiesburg, MS 39406, USA*

Despite enormous efforts, our understanding the structure and dynamics of α-synuclein (ASN), a disordered protein (that plays a key role in neurodegenerative disease) is far from complete. In order to better understand sequence-structure-property relationships in α-SYNUCLEIN we have developed a coarse-grained model using knowledge-based residue-residue interactions and used it to study the structure of free ASN as a function of temperature (T) with a large-scale Monte Carlo simulation. Snapshots of the simulation and contour contact maps show changes in structure formation due to self-assembly as a function of temperature. Variations in the residue mobility profiles reveal clear distinction among three segments along the protein sequence. The N-terminal (1-60) and C-terminal (96-140) regions contain the least mobile residues, which are separated by the higher mobility non-amyloid component (NAC) (61-95). Our analysis of the intra-protein contact profile shows a higher frequency of residue aggregation (clumping) in the N-terminal region relative to that in the C-terminal region, with little or no aggregation in the NAC region. The radius of gyration (R_g) of ASN decays monotonically with decreasing the temperature, consistent with the finding of Allison et al. (JACS, 2009). Our analysis of the structure function provides an insight into the mass (N) distribution of ASN, and the dimensionality (D) of the structure as a function of temperature. We find that the globular structure with $D \approx 3$ at low T, a random coil, $D \approx 2$ at high T and in between ($2 \leq D \leq 3$) at the intermediate temperatures. The magnitudes of D are in agreement with experimental estimates (J. Biological Chem 2002). © *2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1063/1.4927544]

I. INTRODUCTION

α-synuclein (ASN)[1] is a 140 amino acid protein that is abundant in neurons and shows extensive interactions with the phospholipid membrane and other proteins. ASN has been identified as a critical component in the onset of neurodegenerative diseases (synucleinopathies[2]), including Parkinson's disease (PD), Lewy body dementia and Alzheimer disease. ASN[1,3] has been extensively studied as an intrinsically disordered (unstructured) protein with a monomeric conformation comparable to a random coil. The self-association and toxic clumping of ASN into amyloid fibrils is one of the prominent pathological characteristics that lead to PD containing secondary structures.[4,5] Recent studies[6–11] have shown, however, that ASN can assume a number of structures involving α-helices, β-sheets, trimers and tetramers that resist aggregation. The primary structure of 140 residues ASN[1] consists of three domains, (i) an N-terminal region (residues 1-60) with the propensity to form an alpha helix on membrane binding, (ii) a central region (residues 61-95) with a non-amyloid component (NAC), and (iii) an acidic C-terminal region (residues 96-140).

ASN has been extensively studied, and the contradictory results[6–11] on the structure of ASN has continued to attract enormous interest in this field.[12–21] For example, Mysling et al.[22] has studied the

backbone dynamics of soluble ASN oligomers using hydrogen/deuterium exchange and found that the C-terminal region (residues 94-140) and N-terminal (residues 4-17) domains are very mobile. Gurry et al.[23] used NMR and SAX methods to investigate the ensemble structures. They[23] find that the large fraction of ensemble is a disordered monomer with a small fraction of trimeric and tetrameric oligomers

Enormous efforts[24–34] have been made in understanding the structure and dynamics of ASN using NMR data, molecular dynamics (MD), and Monte Carlo simulations. These studies enhance our understanding of sequence-function correlations, but also reveal significant gaps in our knowledge. Coskuner and Wise-Scira[30] have acknowledged the 'valuable insight' gained from the experimental studies and pointed out that the 'atomic level information with dynamics can be gained from theoretical studies of ASN and its mutation at the monomeric level in solution that are not easily observable using conventional experimental tools. They find that the A53T mutant-type ASN structures are thermodynamically more stable than those of the wild-type protein in aqueous solution with higher propensity to aggregate due to increased β-sheet formation and lack of 'strong intramolecular long-range interaction.' Jonsson et al.[34] have carried out Monte Carlo studies of free ASN and identified both the disordered phase, and phase stabilized by β-strand formation. The importance of all-atom MD simulations[31–33] and its ability 'to capture experimentally observed features' have also been reported.[34] These studies have shown that 'it remains a challenge to explore the full conformational ensemble populated by a flexible protein of this length' and justified 'Monte Carlo (MC) rather than MD methods' involving efficient global moves, e.g., 'pivot update'. Global moves such as 'pivot update'[35] adopted and emphasized appropriately in this study[34] appear 'much more efficient, compared to "small steps" algorithms like MD'. Efficiencies and pitfalls of both MC and MD have been extensively explored in modeling polymers.[36] In these studies we implement the well-tested and efficient procedures, the bond fluctuation scheme,[36,37] in modeling the structure and dynamics of un-solvated ASN.

It is not computationally feasible to incorporate all atomic-scale details to explore the complete conformational phase space of such a protein as large as ASN using the force-fields generally adopted in MD simulations. Coarse grained methods have been used to carry out large-scale computer simulations and draw meaningful conclusions about the sequence-structure-function relationships. Devising interaction potentials, exploring the phase space selectively, resorting to efficient and effective methods, etc. are common procedures in coarse-grained modeling.[38–49] Knowledge-based contact matrix[50–57] (derived from an ensemble of frozen structures of protein available at the protein data bank (PDB)) has been extensively used to develop phenomenological residue-residue interactions to understand the folding dynamics of proteins. As in our previous investigations,[56,57] we will use the classical knowledge-based interaction due to Miyazawa and Jernigan (MJ)[51] and one of its improved versions by Betancourt and Thirumalai (BT)[53] to study the structure and dynamics of un-solvated ASN as a function of temperature.

II. MODEL AND METHOD

In our coarse-grained description,[56,57] ASN is represented by a chain of 140 nodes tethered together by fluctuating bonds (^{1}M)—(^{2}D)—(^{3}V)— . . . —(^{70}V)— . . . —(^{140}A) where each node represents an amino acid. The intra-molecular details of the amino acids are thus ignored but the specificity is captured via its unique interaction energy. The protein chain is placed on a cubic lattice in a random configuration at the start of the simulation, and the bond length between consecutive nodes varies between 2 and $\sqrt{10}$ in units of lattice constant.[36] Despite the simple matrix grid, this approach provides ample degrees of freedom for each residue to move and peptide bonds to fluctuate, much more than that with the fixed bond length frequently used in lattice simulations.[36] Small step (one lattice constant) moves retain some of the small scale details, which may be missed in pivot updates and other arbitrary moves.[36] Because of the efficiency and effectiveness, such a bond-fluctuating mechanism has become a common tool in computer simulation modeling of complex systems as is the case for homopolymers,[36] proteins,[56,57] membranes,[58] and bio-functionalized nano assemblies.[59,60] Each residue interacts[56,57] with the neighboring residues within a range (r_c)

with a generalized Lennard-Jones potential,

$$U_{ij} = \left[|\varepsilon_{ij}| \left(\frac{\sigma}{r_{ij}} \right)^{12} + \varepsilon_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], \quad r_{ij} < r_c \tag{1}$$

where r_{ij} is the distance between the residues at site i and j; $r_c = \sqrt{8}$ and $\sigma = 1$ in units of lattice constant. The potential strength, ε_{ij}, is unique for each interaction pair with appropriate positive (repulsive) and negative (attractive) values used from the knowledge-based contact interactions MJ[51] and BT.[53] The number of interacting lattice sites (within the range of the interaction) of a residue is relatively large (on the order of a hundred). Because of the efficiency of the approach with the fluctuating covalent bond, it is easier to explore the huge conformational space while incorporating ample degrees of freedom.[36,37]

Each tethered residue performs its stochastic movements with the Metropolis algorithm briefly described as follows. A residue at a site i is selected randomly to move to a neighboring lattice site, j. The excluded volume constraints are then checked, including the covalent bond length as a result of the proposed random move. If satisfied, the residue is moved from site i to site j with the Boltzmann probability $exp(-\Delta E_{ij}/T)$, where $\Delta E_{ij} = E_j - E_i$ is the change in energy between its new (E_j) and old (E_i) configuration; T is the temperature in reduced units of the Boltzmann constant and the energy (ε_{ij}), and an attempt to move each residue once defines the unit Monte Carlo step (MCS).[35] We monitor a number of local and global physical quantities during the course of simulation, including the energy of each residue, its mobility, mean square displacement of the center of mass of the protein, radius of gyration and its structure factor. Simulations are performed at each temperature for a sufficiently long time (typically ten million time steps) with many independent samples (typically 100 samples) to estimate the average values of these quantities. We have used a 64^3 lattice to generate all the data presented here although different lattice sizes are also used to verify that our findings are independent of the finite size qualitatively.

III. RESULTS AND DISCUSSION

All physical quantities and variables for ASN are presented in arbitrary (reduced) units as noted above. The simulation temperature is varied to assess the variations in these physical quantities, and we focus in a range where most changes in these physical quantities occur by avoiding the low and high temperature extremes.

Typical snapshots from the simulations as a function of temperature are presented in Figure 1. Although a snapshot does not provide a comprehensive summary of the average ensemble behavior (involving millions of configurations), it provides a glimpse into some of the characteristics. At low temperatures, globular structures (multi-scale segmental scales to overall global) appear. Raising the temperature opens up the compact structures resulting in loop formation, and fibrous structures prevail at high temperatures where localized residue aggregation and small loops persist.

Contour maps of the snapshot configurations are presented in Figure 2 for a representative set of temperatures ranging from 0.026 to 0.032. These results show a systematic reduction in looping
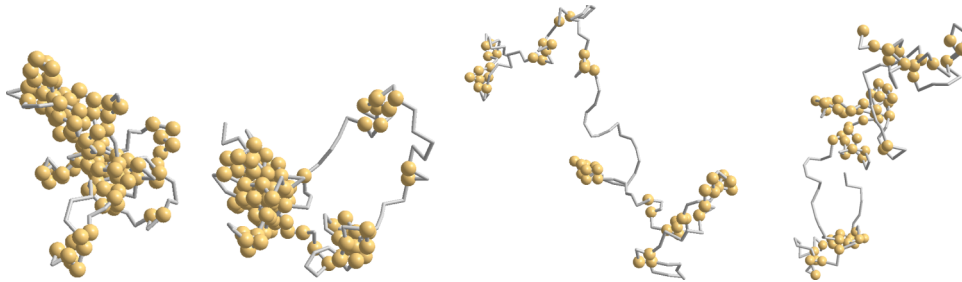

FIG. 1. Snapshots of protein configurations at the end of 10 million time steps at temperatures T = 0.026, 0.028, 0.030, 0.032 (from left to right). The nodes (residues) close enough to experience interactions are shown by spheres.
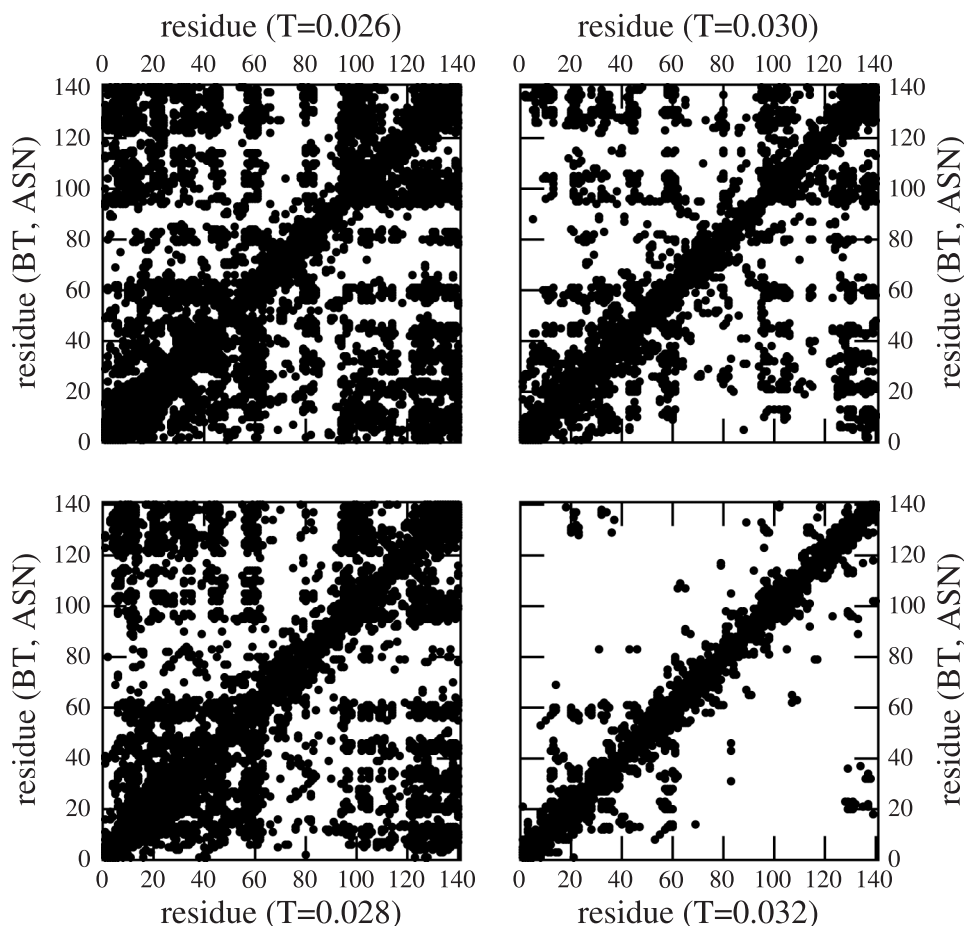
FIG. 2. Residue map (neighboring residues along the contour within the range of interaction) of protein configurations at the end of 10^7 time steps with 40 independent runs at temperatures T = 0.026, 0.028, 0.030, and 0.032 with BT potential.[53]

with increasing the temperature. Although it is difficult to compare these trends quantitatively with the results from different models, the general features in distribution of contact loops (at T = 0.030) appear similar to of the experimental results of Dedmon et al. (Figure 2).[31]

The mobility of each ASN residue in the simulation as a function of temperature can be used to identify the more mobile segments. The average residue mobility (M_n) is defined as the fraction of its successful moves per unit time step, and Figure 3 shows the mobility profile of ASN at temperatures T = 0.026, 0.028, 0.030, and 0.032. The least mobile residues at T = 0.026 include ^{13}E, ^{19}A, ^{21}K, ^{23}K, ^{28}E, ^{32}K, ^{34}K, ^{35}E, ^{43}K, ^{45}K, ^{46}E, ^{56}A, ^{57}E, ^{58}E, ^{60}K in the N-terminal region and ^{96}K, ^{97}K, ^{102}K, ^{104}E, ^{105}E, ^{129}S, ^{131}E in the C-terminal region. The intermediate (61-95) comprising the non-amyloid component (NAC) is more mobile relative to the N-terminal and C-terminal regions. The least mobile residues are predominantly E, K, and A, and such rigidity may provide seeds for aggregation. Aggregation of the N-terminal region of the protein may be enhanced by the close proximity of attractive residues compared to the C-terminal region. Apart from the steric constraints imposed by peptide bonds, the residue mobility depends on the local structure. The data shows that looping may involve residues that are well separated in the linear sequence, especially at low temperatures (Figure 2). Raising the temperature enhances the mobility, and at the highest temperature (T = 0.032) all residues become highly mobile as the residue-residue interactions become less important and the protein assumes a self-avoiding walk (SAW) or random coil conformation (vida infra).

The simulations show that the local structure and mobility are correlated, especially at low temperatures. We quantitate the local structure by considering the average number (N_n) of surrounding residues within the range of interaction as a function of temperature as shown in Figure 4.
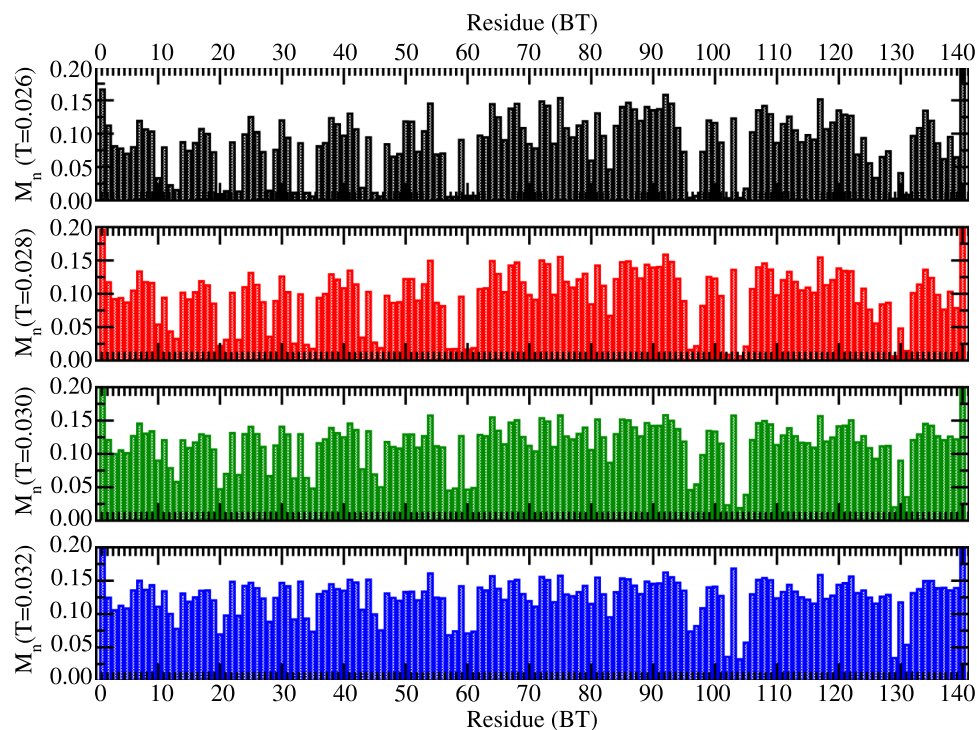
FIG. 3. Average mobility (fraction of successful moves per unit time step) of each residue of ASN protein at temperatures T = 0.026-0.032. Simulations are performed on a 64^3 lattice for 10^7 time steps with 100 independent samples at each temperature with BT potential;[53] the estimates of mobility involve about a billion moves for each residue.
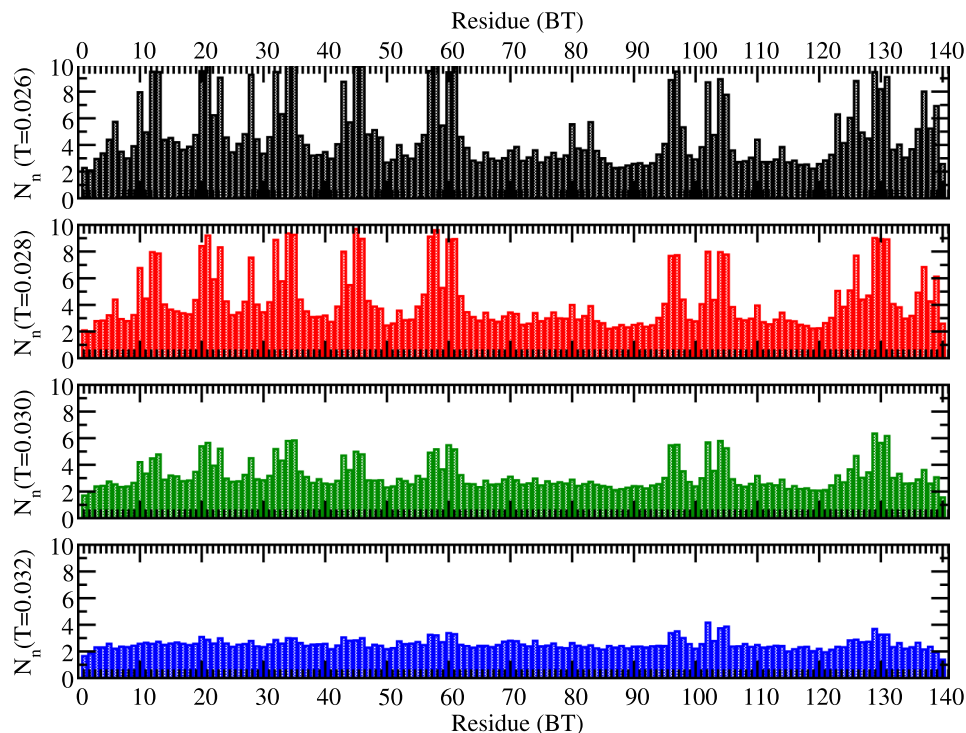

FIG. 4. Average number (N_n) of surrounding residues within the range of interaction of each residue of ASN protein at temperatures T = 0.026-0.032. Statistics is the same as figure 3.

The peaks observed at lower temperatures (T = 0.026, 0.028) are signatures of intra-chain self-organization (either protein folding or aggregation) of the underlying residues. The largest fraction of interacting residues are located in the N-terminal region (1-61) and in the C-terminal region (96-140). Lack of organization in the NAC region (61-95) makes it possible for ASN to assemble in a fibrous structure. Structure formation in the N-terminal region involve ^{10}K, ^{12}K, ^{13}E, ^{20}E, ^{21}K, ^{23}K, ^{27}A, ^{32}K, ^{34}K, ^{35}E, ^{43}K, ^{45}K, ^{46}E, ^{57}E, ^{58}K, ^{60}K, ^{61}E and ^{96}K, ^{97}K, ^{102}K, ^{104}E, ^{105}E, ^{126}E, ^{129}S, ^{130}E, ^{131}E, ^{137}E in the C-terminal region. The residue-residue interactions in the N-terminal and C-terminal regions of the protein suggests the important role N and C-terminals play in controlling the multi-scale structure of the protein. The formation of ASN fibers have been associated with intermolecular β-sheet formation, which cannot be directly identified from our coarse-grained model. However, we note that the spacing and frequency of the interacting residues (primarily due to the location of E and K residues) are consistent with the general structural patterns reported in Ref. 34.

Stochastic movement of residues and their transient settling lead to global motion of the protein that depends on the temperature. The global motion can be examined by analyzing the variation of the root mean square displacement (RMS) of the center of mass of the protein (R_c) as a function of time (t). Figure 5 shows the RMS displacement R_c as a function of time over a range of temperatures (T = 0.026-0.034). The asymptotic variation shows the power-law dependence of the RMS displacement on the time step (t), with $R_c \propto t^\nu$. The power-law exponent ν provides an insight into the type of motion, including diffusion ($\nu = 1/2$), sub-diffusion ($\nu < 1/2$), and drift dynamics ($\nu = 1$). Figure 5 shows a systematic change in the magnitude of the exponent from diffusive $\nu = 1/2$ motion at high temperature (T = 0.034) to sub-diffusive $\nu < 1/2$ dynamics at low temperatures. The power-law exponent ν is so low at lower temperatures (T = 0.026, 0.027) that the protein is essentially stationary in our simulations. We have also analyzed the RMS displacement of the center node (^{70}V) which follows the global dynamics of the protein in the long time.

The protein size depends on the residue-residue interactions and the temperature, and is characterized by the radius of gyration (R_g). We have analyzed the radius of gyration in detail as a function
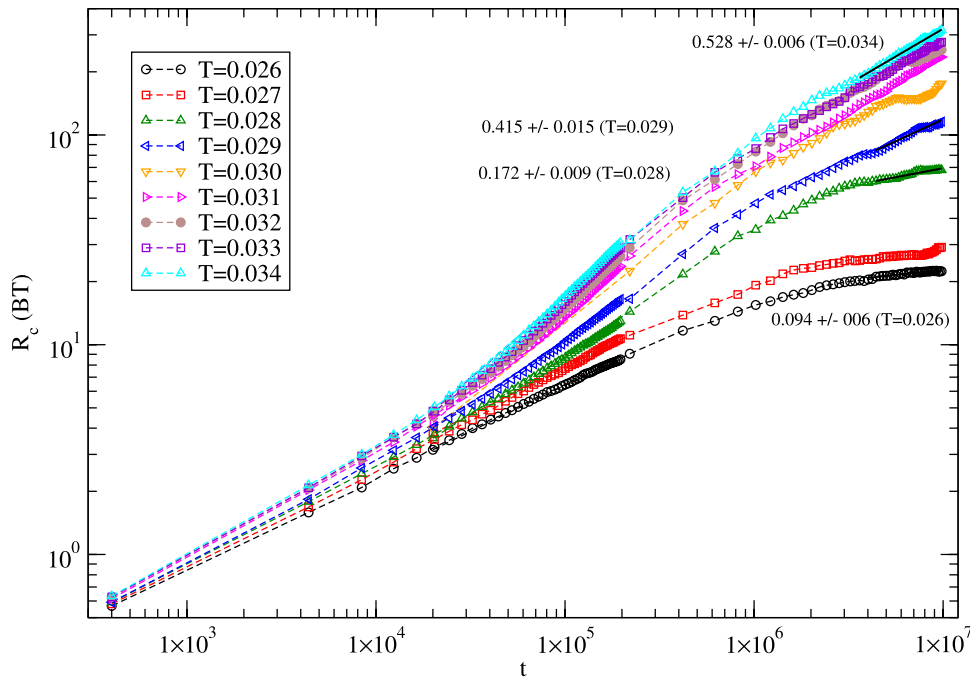

FIG. 5. Variation of the root mean square (RMS) displacement of the center of mass of the protein with the time steps. The numbers along the final data points are the value of the exponent ν in $R_c \propto t^\nu$ ($\nu = 1/2$ shows diffusion, $\nu < 1/2$ is sub-diffusion). Simulations are performed on a 64^3 lattice for 10^7 time steps with 100-1000 independent samples at each temperature with BT potential.[53]
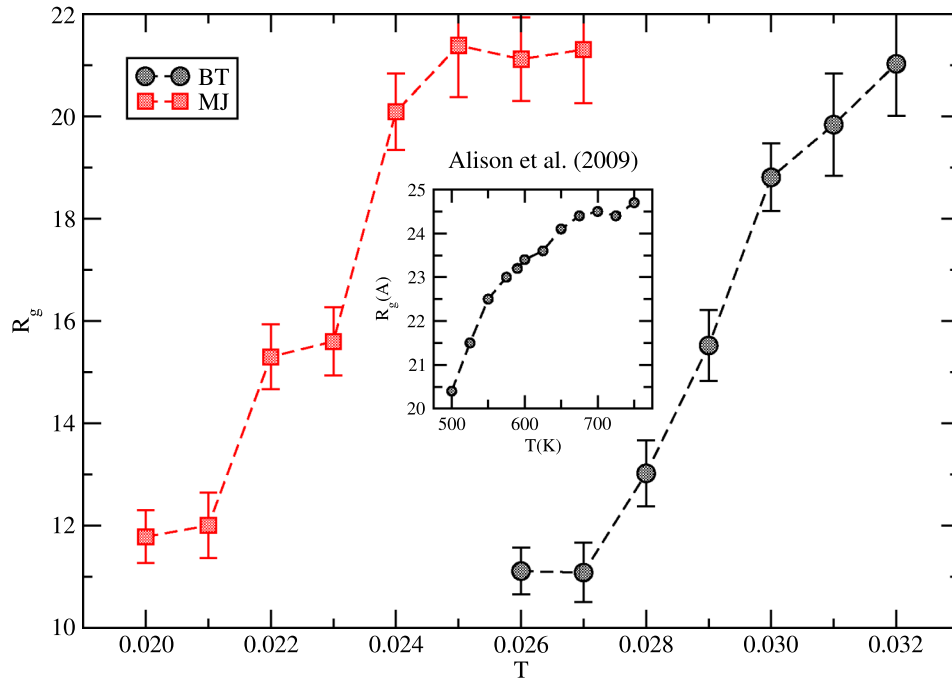
FIG. 6. Variation of the equilibrium average radius of gyration versus temperature with two knowledge-based potentials, classic MJ[51] and improved BT[53] potentials. Inset is the data from Allison et al.[32] Estimates of R_g involve average over the time steps (last one-third to one-half, i.e., the data from last 2.5×10^6 to 1.6×10^6 MCS time) and 100 independent samples each on a 64^3 lattice.

of temperature as the protein relaxes and reaches its equilibrium. R_g was calculated from the last half to one-third of the total time steps (i.e., last 2.5×10^6 to 1.6×10^6 MCS time) and was averaged over a 100 independent samples.

Figure 6 shows the variation of the equilibrium radius of gyration as a function of temperature simulated using two knowledge-based residue-residue interaction potentials (BT, MJ).[51,53] Differences in the BT and MJ potentials leads to a different temperature range for the unfolding transition. While quantitatively different, the variation remain similar qualitatively. A data set is also included from table 3 of Alison et al.[32] which may be the estimate of R_g in a different solvent medium for comparison. This result shows that the qualitative variation of R_g with the temperature are similar to those reported Allison et al.[32]

Structure factor, $S(q)$, of the protein provides an insight into its multi-scale structures. We have studied the structure factor, $S(q)$, as its structures evolve with the temperature,

$$S(q) = \langle \frac{1}{N} \left| \sum_{j=1}^{N} e^{-i\vec{q} \cdot r_j} \right|^2 \rangle_{|\vec{q}|} \tag{2}$$

where r_j is the position of each residue and $|q| = 2\pi/\lambda$ is the wave vector of wavelength, λ. From the power-law scaling of the structure factor with the wave vector, $S(q) \propto q^{-1/\delta}$, one can estimate the spatial distribution of residues in the protein by analyzing its radius of gyration (R_g). The scaling of the radius of gyration of the protein chain with the number N of its nodes (residues), i.e., $R_g \propto N^\gamma$ provides an insight into the shape of the chain and allows us to distinguish between random coil and globular protein conformations. For example, $\gamma = 1/2$ represents a random-coil conformation of the protein. Conversely, one can also estimate the effective dimension (D) of the residue distributions within the radius (R_g) of the protein, i.e., $N \propto R_g{}^D, D = 1/\gamma$. Estimates of these exponents for shape and mass distribution (γ, D) of protein requires evaluation of R_g for a number of different N. Unfortunately, we have only a fixed number (N) of residues in a protein, therefore, scaling of R_g with N is not an option to evaluate the mass distribution (i.e., structure)
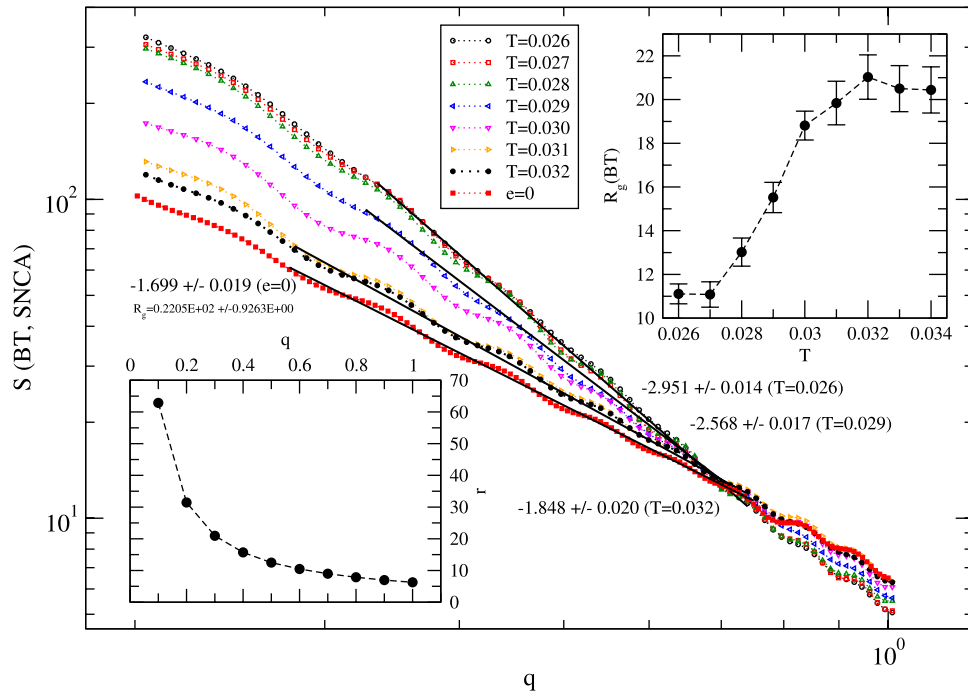
FIG. 7. Variation of the structure factor with the wave vector q for a range of temperatures with BT potential.[53] Slopes of the fitted data points at three representative temperatures are included to show the changes. Insets are the variation of the corresponding radius of gyration with the temperature and that of the wave vector q with the spatial (r) scale. A set of data with only excluded volume interaction (e = 0) is also included for reference, and the theoretical analysis presented in Ref. 31 should be comparable to this reference set. Simulations are performed on a 64^3 lattice for 10^7 time steps with 100 independent samples at each temperature.

of the protein. However, we can estimate the exponents of the mass distribution of protein by analyzing the structure factor over almost all length scales including local segments.

Figure 7 shows the variation, $S(q)$, with the wave vector q with the BT potential. By fitting the data points of comparable proteins ($R_g \approx \lambda$) at appropriate temperatures, we evaluate the effective dimension of ASN. Our data clearly shows a random coil structure ($D \approx 3$) at the low temperature T = 0.026 and random coil ($D \approx 2$), less than 2 if we shift the fitting towards lower q which may be more towards self-avoiding walk rather than random walk at high temperature T = 0.032. These estimates ($\gamma = 1/D$) of γ are consistent with the observations made by Uversky et al. (see their equation 2 and 3).[6] In the absence of residue-residue interaction (e = 0), the above scaling gives $D \approx 1.7$ with corresponding SAW exponent $\gamma = 0.59$. Thus the structure function provides the overall shapes and size of the ASN over the range of temperatures, from low to high.

IV. CONCLUSIONS

A coarse-grained simulation[55,56] with knowledge-based interaction potentials[50,52] is used to investigate the structure of the intrinsically disordered protein ASN as a function of temperature. In their recent study, Jonsson et al.[34] have noted that MC simulations are more effective for exploring the phase space of the conformational ensemble. We have implemented an efficient and effective Monte Carlo scheme with phenomenological residue-residue interactions in order to capture the large scale (both length and temporal) features without sacrificing small scale resolution. The coarse-grained study of free ASN presented here complements the extensive atomic-scale studies while allowing us to explore a larger phase space. We are able to analyze a number of local and global physical quantities and identify the trends in structural evolution of ASN as a function

of temperature. Even though the reduced units (length and time) are arbitrary, these simulations provides an insight into the the trends in the structure and dynamics as a function of temperature.

The contour maps of residue-residue interactions from the MC simulations provides an insight into the structure as a function of temperature and the mobility profiles allows us to identify the most active residue promoting intra- and intermolecular interactions. We find that different segments of the protein, i.e., N-terminal (1-60), C-terminal (96-140), and NAC (61-95) regions, can be clearly distinguished based on the mobile profile, and we are able to identify the least mobile residues in the N-terminal and C-terminal regions around which the local structures may be organized.

We find that the residue-residue contacts that may lead to higher order multimer structures are present at a higher level in the N-terminal region compared the C-terminal, and little globular structure is observed in the in the NAC region. In coordination with the mobility profile, the intra-protein structural profile provides us a mechanism to identify the residues that act as seeds to form local clumps in the N-terminal and C-terminal regions of the ASN. To our knowledge, identification of specific residues with the least mobility and those that can act as seeds for local aggregation are not observed in atomic-scale simulations.

Analysis of the global physical quantities (RMS displacements of ASN, its radius of gyration, and structure factor) provides insight into (*i*) the relaxation and characteristic dynamics, (*ii*) variation of the overall shape, and (*iii*) its scaling with size (distribution of residues) as a function of temperature. We are able to predict the diffusive and sub-diffusive nature of the global dynamics on reducing the temperature from high to low values. Such characteristics will help in understanding how fast ASN can respond at different temperatures. We find that the radius of gyration of free ASN decreases continuously on reducing the temperature, similar to findings reported by Allison et al.[31] Additionally, we are able to identify the random coil conformation at high temperature and globular structures at lower temperatures, which help with understanding how the residues distribute at different temperature. The scaling of the size of ASN with its molecular weight are consistent with the findings reported by Uversky et al.[6] ASN has been extensively studied both experimentally as well as computationally with growing interest to uncover its unknown characteristics that may be related to function and disease states. We hope our findings of the structural response of free ASN to temperature, complement the current simulations, and provide additional tools for the interpretation of the laboratory observations. There are many parameters such as the effects of solvent and the underlying matrix in vivo and in vitro scenarios including protein concentration[61] that we will explore in our future efforts.

ACKNOWLEDGEMENTS

[1] http://en.wikipedia.org/wiki/Alpha-synuclein (December 23, 2013).

[2] A. Bellucci, M. Zaltieri, L. Navarria, J. Grigoletto, C. Missale, and P. Spano, Brain Res. **1476**, 183 (2012).

[3] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury, Biochemistry **35**, 13709 (1996).

[4] M. G. Spillantini, M. L. S., V.M.-Y. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert, Nature **388**, 839 (1997).

[5] V. N. Uversky, J. Li, and A. L. Fink, J. Biol. Chem. **276**, 10737 (2001).

[6] V.N. Uversky, J. Li, P. Souillac, S. Doniac, R. Jakes, M. Goedert, and A.L Fink, J. Biological Chem. **277**, 11970 (2002).

[7] T. Bartels, J. G. Choi, and D. J. Selkoe, Nature **477**, 107 (2011).

[8] W. Wang, I. Perovic, J. Chittuluru, A. Kaganovich, L. T. T. Nguyen, J. Liao, J. R. Auclair, D. Johnson, A. Landeru, A. K. Simorellis, S. Ju, M. R. Cookson, F. J. Asturias, J. N. Agar, B. N. Webb, C. Kang, D. Ringe, G. A. Petsko, T. C. Pochapsky, and Q. Q. Hoang, Proc. Natl. Acad. Sci. **108**, 17797 (2011).

[9] A. J. Trexler and E. Rhoades, Protein Sci. **21**, 601 (2012).

[10] B. Fauvet, M. K. Mbefo, M. B. Fares, C. Desobry, S. Michael, M. T. Ardah, E. Tsika, P. Coune, M. Prudent, N. Lion, D. Eliezer, D. J. Moore, B. Schneider, P. Aebischer, O. M. El-Agnaf, E. Masliah, and H. A. Lashuel, J. Biol. Chem. **287**, 15345 (2012).

[11] A. Binolfi, F. X. Theillet, and P. Selenko, Biochem. Soc. Trans. **40**, 950 (2012).

[12] C. Camilloni and M. Vendruscolo, J. Phys. Chem. **117**, 10737 (2013).

[13] M. Haj-Yahya, B. Fauvet, Y. Herman-Bachinsky, M. Hejjaoui, S. N. Bavikar, S. V. Karthikeyan, A. Ciechanover, H. A. Lashuel, and A. Brik, Proc. Natl. Acad. Sci. **110**, 17726 (2013).

[14] M. M. Ouberai, J. Wang, M. J. Swann, C. Galvagnion, T. Guilliams, C. M. Dobson, and M. E. MWelland, J. Biological Chemistry **288**, 20883 (2013).

[15] E. Hellstrand, A. Nowacka, D. Topgaard, S. Linse, and E. Sparr, PLoS One **8**, e77235 (2013).

[16] T. R. Alderson and J. L. Markley, Intrinsically Disordered Proteins **1**, e26255 (2013).

[17] A. S. Maltsev, J. Ying, and A. Bax, Biochemistry **51**, 5004 (2012).

[18] D. C. DeWitt and E. Rhoades, Biochemistry **52**, 2385 (2013).

[19] Z. Jiang, M. de Messieres, and J. C. Lee, J. Am. Chem. Soc. **135**, 15970 (2013).

[20] H. Okazaki, Y. Ohori, M. Komoto, Y.-H. Lee, Y. Goto, N. Tochio, and C. Nishimura, FEBS Letters **587**, 3709 (2013).

[21] L. Bousset, L. Pieri, G. Ruiz-Arlandis, J. Gath, P. H. Jensen, B. Habenstein, K. Madiona, V. Olieric, A. Bockmann, B. H. Meier, and R. Melki, Nature Communication DOI:10.1038/ncomms3575.

[22] S. Mysling, C. Betzer, P. H. Jensen, and T. J. D. Jorgensen, Biochemistry **52**, 9097 (2013).

[23] T. Gurry, O. Ullman, C. K. Fisher, I. Perovic, T. Pochapsky, and C. M. Stultz, J. Am. Chem. Soc. **135**, 3865 (2013).

[24] L. Cruz, B. Urbanc, J. M. Borreguero, N. D. Lazo, D. B. Teplow, and H. E. Stanley, Proc. Natl. Acad. Sci. U. S. A. **102**, 18258 (2005).

[25] Y. C. Xu, J. J. Shen, X. M. Luo, W. L. Zhu, K. X. Chen, J. P. Ma, and H. L. Jiang, Proc. Natl. Acad. Sci. U. S. A. **102**, 5403 (2005).

[26] A. Baumketner and J. E. Shea, J. Mol. Biol. **362**, 567 (2006).

[27] I. Brovchenko, R. R. Burri, A. Krukau, A. Oleinikova, and R. Winter, J. Chem. Phys. **129**, 195101 (2008).

[28] L. Triguero, R. Singh, and R. Prabhakar, J. Phys. Chem. B **112**, 7123–7131 (2008).

[29] C. Lee and S. Ham, J. Comput. Chem. **32**, 349–355 (2011).

[30] O. Coskuner and O. Wise-Scira, ACS Chem. Neurosci. **4**, 1101 (2013).

[31] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, J. Am. Chem. Soc. **127**, 476 (2005).

[32] J. R. Allison, P. Varnai, C. M. Dobson, and M. Vendruscolo, J. Am. Chem. Soc. **131**, 18314 (2009).

[33] K.-P. Wu, D. S. Weinstock, C. Narayanan, R. M. Levy, and J. Baum, J. Mol. Biol. **391**, 784 (2009).

[34] S. A. Jonsson, S. Mohanty, and A. Irback, Proteins **80**, 2169 (2012).

[35] M. Lal, Mol Phys **17**, 57 (1969).

[36] *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, edited by K Binder (Oxford University Press, 1995).

[37] I. Carmesin and K. Kremer, Macromolecules **21**, 2819 (1988).

[38] A.P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).

[39] J. Zhou, S. Chen, and S. Jiang, Langmuir **19**, 3472-3478 (2003).

[40] A. E. van Giessen and J. E. Straub, J. Chem. Phys. **122**, 0249041 (2005).

[41] D. Reith, M. Putz, and F. Muller-Plathe, J. Comput. Chem. **24**, 1624 (2003).

[42] R. B. Pandey and B. L. Farmer, J. Chem. Phys. **132**, 125101 (2010).

[43] A. Liwo, C. Czaplewski, S. Oldziej, and H. A. Scheraga, Curr. Opin. Struct. Biol. **18**, 134 (2008).

[44] F. Ercolessi and J. Adams, Europhys. Lett. **26**, 583 (1994).

[45] J. Zhou, I. F. Thorpe, S. Izvekov, and G.A. Voth, Biophys. J. **92**, 4289 (2007).

[46] D. H. de Jong, G. Singh, W. F. Drew Bennett, C. Arnarez, T. A. Wassenar *et al.*, J. Chem. Theory Comput. **9**, 687 (2013).

[47] J. Sorensen, P. Xavier, K. K. Skeby, S. J. Marrink, and B. Schiott, J. Phys. Chem. Lett. **2**, 2385 (2011).

[48] T. Haliloglu and I. Bahar, Proteins **31**, 271 (1998).

[49] Z. Wu, Q. Cui, and A. Yethiraj, J. Chem. Theory and Compu. **7**, 3793 (2011).

[50] S. Tanaka and H. A. Scheraga, Macromolecules **9**, 945 (1976).

[51] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).

[52] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 623 (1996).

[53] M. R. Betancourt and D. Thirumalai, Protein Sci **2**, 361 (1999).

[54] A. Godzik, A. Kolinski, and J. Skolnick, Proteins **4**, 363 (1996).

[55] S.-Y. Huang and Z. Xiaoqin, Proteins **79**, 2648 (2011).

[56] M. Fritsche, R. B. Pandey, B. L. Farmer, and D. Heermann, PLoS one **7**, e32075 (2012).

[57] R. B. Pandey and B. L. Farmer, PLoS one **7**, e49352 (2012).

[58] R.B. Pandey *et al.*, J. Polym. Sci. B: Polym. Phys. **48**, 2566 (2010).

[59] R.B. Pandey *et al.*, Phys. Chem. Chem. Phys. **11**, 1989 (2009).

[60] R. B. Pandey and B. L. Farmer, J. Chem. Phys. **139**, 164901 (2013).

[61] R. B. Pandey, B. L Farmer, and B.S. Gerstman, AIP Advances **5**, 092502 (2015).